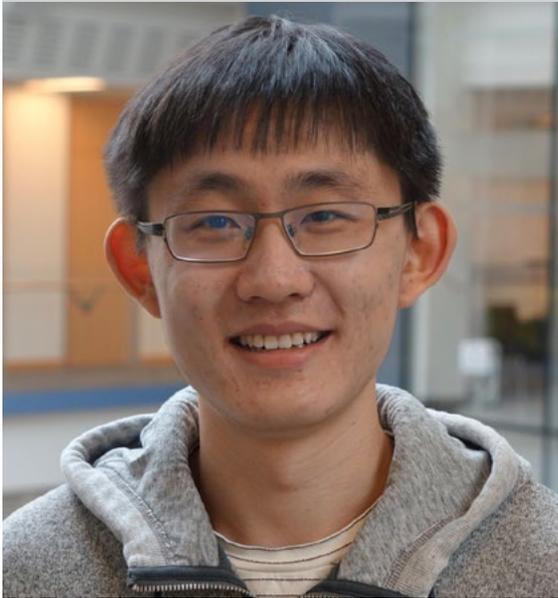




## DISSERTATION DEFENSE



# ZEYU ZHENG

## Advances in Deep Reinforcement Learning: Intrinsic Rewards, Temporal Credit Assignment, State Representations, and Value-equivalent Models

Thursday, June 16, 2022

9:00 – 11:00am

3725 Beyster

Hybrid – [Zoom](#)

**ABSTRACT:** Reinforcement learning (RL) is a machine learning paradigm concerned with how an agent learns to predict and to control its own experience so as to maximize long-term cumulative reward. In the past decade, deep reinforcement learning (DeepRL) emerged as a new subfield that aims to combine sequential decision-making techniques in RL with powerful function approximation tools offered by deep learning. DeepRL has seen great success such as defeating human champions in Go and has made an impact on real-world applications like robot control.

This thesis aims to further advance DeepRL techniques. To this end, we identify several challenges that current DeepRL agents face and propose methods to address them. Concretely, we address the following four challenges: 1) The challenge of designing reward functions that helps agents learn efficiently. We propose a novel meta-learning algorithm for learning reward functions that facilitate policy optimization. Our algorithm improves the performance of policy-gradient methods and outperforms handcrafted heuristic reward functions. In a follow-up study, we show that the learned reward functions can capture knowledge about long-term exploration and exploitation and can generalize to different RL algorithms and changes in the environment dynamics. 2) The challenge of effective temporal credit assignment. We explore methods based on pairwise weights that are functions of the state in which the action was taken, the state in which the reward was received, and the time elapsed in between. We develop a metagradient algorithm for adapting these weights during policy learning. Our experiments show that our method achieves better performance than competing approaches. 3) The challenge of learning good state representations. We investigate using random deep action-conditional predictions tasks as auxiliary tasks to help agents learn better state representations. Our experiments show that random deep action-conditional predictions can often yield better performance than handcrafted auxiliary tasks. 4) The challenge of learning accurate models for planning. We propose a new method for learning value-equivalent models, a class of models that demonstrates strong empirical performance lately, that generalizes existing methods. Our experiments show that our method can improve both the model prediction accuracy and the control performance of the downstream planning procedure.

**CHAIR:** Prof. Satinder Singh Baveja